AGU100 ADVANCING EARTH AND SPACE SCIENCE

Earth and Space Science

RESEARCH ARTICLE
10.1029/2019EA000794

Key Points:
- The frequency of synoptic wind patterns are used to assess the consistency between reanalyses
- ERA-Interim, ERA5, MERRA2, and JRA55 show high consistency around Antarctica in the satellite era using the new entropy coefficient metric
- CERA20C, ERA20C, and 20CRV2c perform poorly before 1957 as measured by the entropy coefficient and ensemble spread and are better after 1957

Correspondence to:
A. J. McDonald,
adrian.mcdonald@canterbury.ac.nz

Citation:
McDonald, A. J., Cairns, L. H. (2020). A new method to evaluate reanalyses using synoptic patterns: An example application in the Ross Sea/Ross Ice Shelf region. *Earth and Space Science*. 7, e2019EA000794. https://doi.org/10.1029/2019EA000794

Received 12 JUL 2019
Accepted 8 NOV 2019
Accepted article online 22 NOV 2019

A New Method to Evaluate Reanalyses Using Synoptic Patterns: An Example Application in the Ross Sea/Ross Ice Shelf Region

Adrian J. McDonald[1,2] **and Luke H. Cairns**[2]

[1]Gateway Antarctica, University of Canterbury, Christchurch, New Zealand, [2]School of Physical and Chemical Sciences, University of Canterbury, Christchurch, New Zealand

Abstract We compare the consistency between eight reanalyses: CERA20C, ERA5, ERA-Interim, ERA20C, NCEP-DOE, MERRA2, JRA55, and 20CRV2c. This comparison uses daily surface winds near Antarctica to classify synoptic patterns using the self-organizing map technique. The relative frequency of occurrence (RFO) of these patterns are very similar during the satellite era in each reanalysis. The three most common patterns are the same in each reanalysis and changes between the reanalyses only display a 12% relative variation. Examination of the RFOs over time highlights that the CERA20C and 20CRV2c reanalyses display large changes previous to 1957. These changes are likely connected to model relaxation toward their climatology because of a lack of observational constraints. Primarily, we introduce the entropy coefficient (U) which quantifies the consistency between reanalyses in their representation of synoptic patterns. Examination of U shows current reanalyses (ERA5, ERA-Interim, JRA55, and MERRA2) are highly consistent in the satellite era likely due to good observational coverage. However, centennial reanalyses (CERA20C, ERA20C, and 20CRV2c) show two upward step changes in consistency as measured by U at around 1957 and to a lesser extent 1979. Low values of U before 1957 suggest that centennial reanalyses are of limited use before this date, but may be useful after 1957 in this region. We also show that the entropy coefficient displays an inverse relationship with ensemble spread metrics of individual reanalyses. We conclude that the entropy coefficient provides a powerful quantification of the influence of changes in observation density on reanalysis quality in data sparse regions.

1. Introduction

Meteorological reanalysis combines in situ and satellite observations with forecasts from a fixed version of a numerical weather prediction system and arguably provide the best estimate of the atmospheric state. Reanalyses also complement observational data sets by describing changes through the historical record. They are also useful because they provide gridded output and physical consistency across variables within the limitations of the model. However, the errors and uncertainties within the observations and models, and correspondingly the reanalyses, are only partially understood. This means that it is unclear how accurate we can expect reanalysis results to be.

This study compares the consistency between eight reanalysis products using a new technique based on a self-organizing map (SOM) clustering scheme of synoptic states. Given that the uncertainty in most reanalyses is largest in regions with few or no observations we test our methodology in a region over the Antarctic continent and the nearby Southern Ocean. While a number of studies have examined the quality of reanalyses in the polar regions, Jakobson et al. (2012) identified that limited quantities of independent observations are available to test the quality of reanalyses in these regions. For example, they indicated that the tethersonde observations they used to compare with reanalyses should not be considered as representing the climatology of the study period or region. This is a common theme in a number of evaluation studies, with the work in Nygard et al. (2016) also suggesting that it is very challenging to make an objective ranking of the reanalyses around the Antarctic at least partially because of the lack of independent data sets.

Methodologies which allow recommendations about the usage of reanalyses are extremely valuable in the Antarctic because in situ observations are sparse and unevenly distributed (Lazzara et al., 2012). Given this sparsity of observations reanalysis products are therefore potentially a particularly valuable tool for studying weather and climate in this region. In addition, despite the availability of nearly four decades

of satellite-based observations, the length of the observational record remains a fundamental limitation to statistical analysis for research focused on climate variability and trends around Antarctica (Jones et al., 2016).

We focus in this paper on the representation of synoptic-scale patterns in surface wind fields over the Ross Ice Shelf/Ross Sea to demonstrate our method. Surface winds are important in helping modulate meridional heat fluxes around the entire continent and have been used in many synoptic climatology studies. Both Hosking et al. (2013) and Coggins and McDonald (2015) have also identified that changes in surface winds connected to the Amundsen Sea Low (Raphael et al., 2016) have an important impact in this area. Surface winds also force the drift of sea ice and define the surface wind stress which plays a key role in the energy and water cycle of the atmosphere-ocean coupled system (Wen et al., 2019). Work in Barthelemy et al. (2018) tested how biases in Antarctic sea ice simulations relate to atmospheric forcing uncertainties using an ocean-sea ice model driven by three different atmospheric reanalyses. They identified that the three products differed substantially in their surface winds from one another in the Southern Ocean, with their results showing that the atmospheric forcing has a strong influence on the speed of the summertime sea ice retreat, the minimum ice extent, the spatial distribution of ice at the minimum extent, and the total sea ice volume. Wen et al. (2019) also identified that surface wind stress has a dominant role in driving ocean circulations, modulating heat and moisture fluxes at the air-sea interface and in the generation of marine waves. The accuracy of surface winds also determines the quality of the ocean state estimations. Li et al. (2013) also identified that accurate surface winds were particularly important because they drive the Antarctic Circumpolar Current eastward around the Antarctic continent. These winds effectively push the surface waters away from the Antarctic continent through Ekman transport, creating a significant divergence-driven upwelling south of the Antarctic Circumpolar Current.

We examine surface winds in the vicinity of the Ross Sea/Ross Ice Shelf (RS/RIS), a region which is challenging to accurately represent because of the low density and reliability of observations in the region. In an effort to gain an insight into the quality of surface winds across the RS/RIS, the frequency of occurrence of representative synoptic weather patterns, derived using the SOM technique applied to ERA-Interim data, is examined across eight reanalyses. The current study builds on previous synoptic climatology research in the region which have characterized the surface wind field (Coggins et al., 2014; Cohen et al., 2013; Seefeldt & Cassano, 2012; Seefeldt et al., 2007), used SOMs to examine the underlying forcing mechanisms of common synoptic features (Coggins & McDonald, 2015; Nigro & Cassano, 2014), showed how synoptic states effect the cloud climatology (Jolly et al., 2018; Silber et al., 2019), and compared surface observations with mesoscale model output (Jolly et al., 2016). This methodology allows us to assess the consistency and the likely relative quality of the reanalyses. This enables us to make recommendations without relying on the limited number of observations not assimilated into the various reanalyses. To our knowledge only the study by Stryhal and Huth (2017) has previously utilized classifications of circulation patterns to evaluate reanalysis data, though analyses of cyclones (Hoskins & Hodges, 2005) and anticyclones (Parsons et al., 2016) are relatively common place. Stryhal and Huth (2017) examined the frequency of occurrence of representative circulation patterns in five reanalyses over a number of European domains. We further develop these techniques to provide a single measure of consistency based on information theory.

A short summary of previous research on various aspects of the reanalysis products over the polar regions is now presented, these studies have generally shown that the reanalyses are limited in these regions, with their quality being particularly poor previous to the beginning of significant quantities of satellite observations circa 1979. Studies have also shown that the quality of reanalyses are hard to quantify at the poles because of the lack of independent data sets and the low density of the observing networks. Nevertheless, studies have displayed poor performance in both the Arctic (Lindsay et al., 2014; Liu & Key, 2016; Marshall et al., 2018; Rapaic et al., 2015; Simmons & Poli, 2015; Screen & Simmonds, 2011) and the Antarctic (Bracegirdle, 2013; Bracegirdle & Marshall, 2012; Bromwich & Fogt, 2004; Bromwich et al., 2007; Fogt et al., 2017; Nygard et al., 2016; Schneider & Fogt, 2018).

In terms of specific weaknesses a number of studies have identified spurious trends in reanalyses in the temperature (Huai et al., 2019; Wang et al., 2016; Zhang et al., 2018), pressure (Bromwich & Fogt, 2004; Schneider & Fogt, 2018), precipitation (Bromwich et al., 2011; Nicolas & Bromwich, 2011), and surface wind (Bromwich et al., 2007; Nygard et al., 2016) fields around the Antarctic. Wohland et al. (2019) also questioned the quality of reanalyses surface wind trends in centennial reanalyses and Thorne and Vose (2010) have

questioned the reliability of long-term trends in atmospheric reanalyses more widely. Nicolas and Bromwich (2011) also suggested that unrealistic trends in NCEP-DOE and JRA-25 in precipitation fields around coastal East Antarctica are caused by spuriously enhanced meridional wind flows, highlighting the importance of the correct simulation of circulation patterns due to temperature advection.

A limited number of studies have focused on winds around Antarctica. Sanz Rodrigo et al. (2013) indicates that previous research shows that winds around Antarctica are controlled by either large-scale (synoptic) or local katabatic (boundary layer) forcing. They examined ERA40 and ERA-Interim relative to the Automated Weather Station (AWS) database (Lazzara et al., 2012) and found a significant improvement in the representation of the near-surface winds in higher horizontal resolution reanalyses, such as ERA-Interim. Though they also see a lower wind variability in the reanalyses relative to AWS data which they attribute to smoothed topography. Nygard et al. (2016) examined eight reanalyses around the Antarctic Peninsula relative to field observations and found that wind speeds were predominantly low biased. Work in Jones et al. (2016) compared AWS wind speeds to four reanalysis products in the Amundsen Sea Embayment. They showed that the reanalyses severely underestimate wind speeds when observations are above 15 m/s and overestimate when observations are below 5 m/s. They therefore suggest caution is required for periods of high wind speeds (>15 m/s) and near complex coastal topography as the reanalyses are unable to adequately capture the variability in winds. Work in Coggins et al. (2014) compared AWS wind speeds with ERA-Interim output over the RS/RIS and found a generally good correlation, but again found that during strong wind events ERA-Interim underestimated wind speeds. A similar result is also being identified near the RIS front in Dale et al. (2017).

Fewer studies have examined winds over the even more sparsely sampled Southern Ocean, but Li et al. (2013) identified that two reanalyses had large positive biases under weak wind regimes, good agreement with the ship observations under moderate wind regimes, and large negative biases under high wind regimes. In addition, Sato et al. (2018) investigated the impact of additional radiosonde observations over the Southern Ocean on the reproducibility and predictability of atmospheric fields in the Southern Hemisphere and found positive results.

Studies have also examined the representation of low-pressure systems in the different reanalyses, with early studies detailed in Hoskins and Hodges (2005) looking at cyclone track characteristics. Work detailed in Bromwich et al. (2007) also showed that cyclonic activity is markedly different between the reanalyses in the Southern Hemisphere polar regions, where there are few matched cyclones prior to 1979. In comparison, they identified that only weaker cyclones are not matched in the Northern Hemisphere. More recently, Befort et al. (2016) compared the representation of extratropical cyclones over both hemispheres in the 20CRV2 and ERA20C reanalysis during the winter seasons. Their results indicate substantial differences between the two data sets, especially in the first half of the twentieth century, suggesting that understanding the differences and similarities between the reanalyses and their representation of synoptic features is an important ongoing effort.

As identified the longer period reanalyses are particularly valuable around the Antarctic. For example, Fogt et al. (2017) and Jones et al. (2016) both identify that the attribution of changes to anthropogenic forcings, such as stratospheric ozone depletion and greenhouse gases, is limited by the length of climate records and the large natural variability around the continent. But previous reanalyses have been shown to be unreliable in Antarctica before the satellite era, due to the lack of adequate data to guide the reanalysis (Bracegirdle & Marshall, 2012; Bromwich & Fogt, 2004; Bromwich et al., 2007). Thus, one of the key motivations for this analysis is to determine whether newer centennial products are more robust or still suffer from this issue. Schneider and Fogt (2018) recently compared three centennial reanalyses previous to the satellite era with a statistically based pressure reconstruction (Fogt et al., 2016), they found that the quality of the pressure field in reanalyses in the high latitudes of the Southern Hemisphere prior to approximately 1950 is compromised by the lack of observations. They also identify that the reanalyses therefore behave like an unguided model in the early period.

2. Data Sets and Methodology

2.1. Reanalyses

This study compares the consistency between eight separate reanalysis products using a new technique based on a SOM clustering scheme. We use the CERA20C (Laloyaux et al., 2018), ERA20C (Poli et al.,

2016), ERA5, ERA-Interim (Dee et al., 2011), JRA55 (Kobayashi et al., 2015), NCEP-DOE (Kanamitsu et al., 2002), MERRA2 (Gelaro et al., 2017), and 20th Century reanalysis Version 2c (20CRV2c; Compo et al., 2011) reanalysis surface winds in the vicinity of the RS/RIS in this study. One subset of this group is the current generation reanalyses that assimilate all observations, this subset includes ERA5, ERA-Interim, JRA55, and MERRA2. Many of these products (ERA5, ERA-Interim, and MERRA2) cover only the period from 1979 onward, which we define as the "satellite era" in this work for simplicity. Among the modern products only JRA55 produces output prior to the satellite era with data starting in 1958. We also examine centennial-scale reanalyses, which examine longer periods and attempt to overcome step changes due to variations in observational networks by limiting the type of observations assimilated. For example, 20CRV2c (Compo et al., 2011), ERA20C (Poli et al., 2016), and CERA20C (Laloyaux et al., 2018) have all been created in this way, with the latter assimilating both atmospheric and oceanic observations. Ensembles of 20CRV2c and CERA20C were also available which allow estimates of uncertainty which we use to compare with our methodology later. However, these ensembles do not measure the model uncertainty component which is likely largest in regions with few or no observations.

A brief discussion of each reanalysis is now provided. CERA20C is a global centennial reanalysis that reconstructs the past weather and climate of the Earth system including the atmosphere, ocean, and sea ice from 1901–2010. CERA20C assimilates a limited number of observations using a 4DVAR scheme with surface pressure and marine wind observations as the primary input. The results presented in this work represent the ensemble mean derived from the 10-member ensemble available. Details of CERA20C are discussed in Laloyaux et al. (2018). ERA20C is a global centennial atmospheric reanalysis from 1900–2010 with limited data assimilation again based on 4DVAR (Poli et al., 2016). We use the ensemble mean data in this study. The 20CRv2c data set is the last centennial-scale reanalysis used in this study and provides data from 1851 to 2012. This is again a limited input reanalyses and assimilates only surface and sea level pressure observations using an Ensemble Kalman Filter scheme. An ensemble of 56 integrations of the model are available to produce an estimate of the atmospheric state and also a measure of uncertainty. The 20CRV2c is discussed in Compo et al. (2011).

The first of the current generation reanalyses we discuss is ERA-Interim, which is a full input reanalyses that assimilates new observations as they become available using 4DVAR. A detailed discussion about the ERA–Interim reanalysis is available in Dee et al. (2011). ERA-Interim is used as a reference in this study since it has been evaluated multiple times and has proven to perform reasonably well against observations around the Antarctic (Bracegirdle, 2013; Bracegirdle & Marshall, 2012). ERA-Interim is to be replaced long term by ERA5 which is the latest climate reanalysis produced by the European Centre for Medium range Weather Forecasting (ECMWF). It is planned that ERA5 will eventually cover the period back to January 1950, but at present only data back to 1979 are available. We also use the Japanese 55-year Reanalysis (JRA55), this reanalysis extends for a 55 year period starting from 1958, when regular radiosonde observations became operational globally. Details about JRA55 are detailed in Kobayashi et al. (2015). MERRA2 is the last current generation reanalysis used in this paper. MERRA2 is a NASA reanalysis and is available over the satellite era and uses the Goddard Earth Observing System Data Assimilation System Version 5 (GEOS-5). MERRA2 is an update on MERRA and assimilates observations not available to the previous version. Details of MERRA2 are available in Gelaro et al. (2017). The older NCEP-DOE reanalysis is also included in this study as a useful reference point for comparison with older literature. NCEP-DOE, sometimes referred to as NCEP2, is the second reanalysis from NCEP and covers the satellite era. Details are available in Kanamitsu et al. (2002).

2.2. Data Analysis

The analysis presented uses the SOM technique applied to ERA-Interim output between 1980 and 1999 to derive representative surface wind patterns, output from the other reanalysis are then attributed to one of these patterns using a Euclidean distance metric. SOMs are an iterative unsupervised learning scheme commonly used in clustering (Kohonen, 1990). The learning process adjusts a set of reference vectors based on the differences between the reference vector and each input record. A learning rate determines how the adjustment is related to the difference between the reference vector and the input data. Training then consists of many iterations of reference vector adjustment until stable values are reached. In each iteration, the best matching reference vector is identified, using the Euclidean distance, for each input record and updated to more closely resemble the input data.

The learning rate and width of the kernel are reduced as a function of time such that the SOM evolves more quickly initially. The Euclidean distance is used to identify reference vectors within a certain range of the best matching vector. The vectors that fall within this neighborhood are then updated, though to a lesser degree than the best matching node, this feature of the scheme producing the coherent organization of output. The training process ultimately produces reference vectors that represent a distinct portion of the multidimensional input space. The end result of the SOM training is an objective set of distinct maps (referred to as nodes from this point on) that in combination are representative of the entire data set.

The SOM methodology allows us to derive representative surface wind regimes over the RS/RIS. As previously indicated, we use the ERA-Interim reanalysis because it has been evaluated around Antarctica and shows generally good performance (Bracegirdle, 2013; Bracegirdle & Marshall, 2012). Note that previous work in Fogt et al. (2018) has also used ERA-Interim as the reference standard for similar reasons. One potential problem with this usage is that the synoptic patterns from ERA-Interim in the SOM training period (1980–1999) might not be fully representative of the early twentieth century patterns. However, given the high uncertainty on whether the patterns have changed this is hard to asses and we have thus decided to use ERA-Interim only as used in Fogt et al. (2018). However, application of the SOM technique to the MERRA2 data set creates a very similar set of nodes (not shown) and thus the usage of ERA-Interim as a reference standard appears robust. Rather than apply the SOM technique directly to all the ERA-Interim output between 1980 and 1999 it was necessary to reduce the quantity of the data input into the SOM and reduce variations smaller than the synoptic scale in the input data. This data reduction step required to reduce the size of the input matrix to the SOM was completed by applying empirical orthogonal function analysis to the space-time cube of the surface winds (both zonal and meridional winds) and then applying the SOM technique to the largest Principal Components (PCs) only. In particular, we truncated the set of PCs when the explained variance reached a threshold of 90% of the total variance. We used the implementation of the SOM methodology in the SOMPAK (v 3.1) software (obtained from http://www.cis.hut.fi/research/som&urluscore;pak/) to derive the SOM from the most important PCs. The usage of the Empirical Orthogonal Functions (EOF) analysis requires anomalies as inputs and the climatological mean from each latitude/longitude point for the 1980–1999 reference period was used to derive anomalies for all years.

Before analyzing the various reanalysis products, each reanalysis was projected onto a common geographic grid based on the lowest horizontal resolution data set (20CRV2c with 2° by 2° grid spacing) over the geographic domain (60–90°S, 140–220°E) considered. This reduction in the horizontal resolution of some products should not impact our aim, which is to examine synoptic scale features, but does mean that high-resolution features, which are strongly topographically dependent, such as katabatic winds, are not considered across the different reanalyses. The winds at 2 m output from the MERRA2 output were also transformed to 10 m winds to more directly compare with the ERA-Interim output under the assumption of a logarithmic wind speed profile. We also choose to mask regions with topography above 500 m because the varying spatial resolution of the reanalyses means that the complex near-coastal topography might be poorly captured in some models, also impacting the variability in the winds (Jones et al., 2016). This also ensured that changes in the representation of distinct topographic features is reduced in this analysis. This mask acts to remove all points south of 86°S from the cluster analysis and more than a third of points south of 66°S, mostly over the Trans-Antarctic mountains. We also only used data from 00:00 UTC on each day to reduce the processing requirements for the study. Note that while the six reanalyses evaluated in Tastula et al. (2013) found that diurnal cycles of near-surface variables were often erroneous, the near-surface wind speed displays low diurnal variability in both observations and in all the reanalyses products and thus selecting only the output at 00:00UTC should not impact our results. In addition, for internal consistency the 1980–1999 climatological mean for each reanalysis was calculated to derive anomalies used to identify the best matches with the representative patterns.

3. Results

Figure 1 displays the 12 representative circulation patterns (or nodes) for the wind speed derived from the application of the SOM technique to ERA-Interim output from 1980–1999. The set of nodes from this reference period are used throughout this study for comparison with all other reanalyses; however, using the MERRA2 reanalyses causes little change in the results identified later in this study. A 3 × 4 SOM (three columns and four rows) was selected for this SOM because it minimized quantization error and represented a good balance in terms of representation of the wind patterns over the region. Note that we display the sum

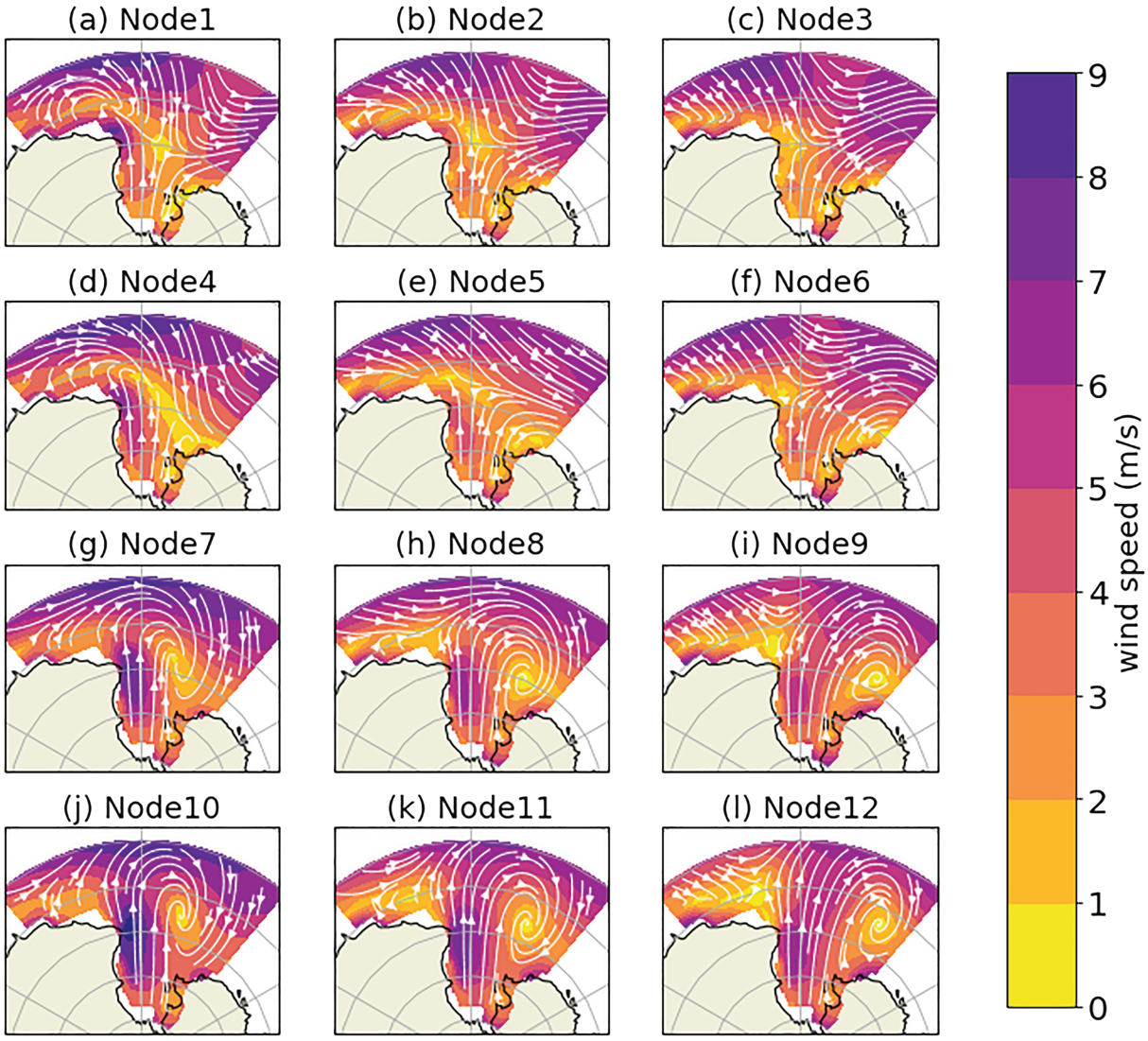

Figure 1. Surface horizontal surface winds for each of the 12 nodes (representative circulation pattern) in the SOM derived from ERA-Interim reanalysis output for the period 1980–1999 are displayed. Streamlines identify the direction of the winds, while the color contours identify the magnitude of the wind speed. Note that the ordering of nodes is identified above each of the representative wind patterns and is useful for reference in later figures.

of the 1980–1999 climatological mean and the anomaly patterns to aid comparison with previous studies. The various nodes in Figure 1 are dominated by strong southerly winds from the interior of the continent over the RIS and westerly winds over the RS in the north of the domain. Each pattern also appears to compare well to clusters in previous classifications. For example, the patterns in Nodes 7, 10, and 11 all show strong low-level jet structures close to the Trans-Antarctic Mountains reminiscent of structures related to the RIS Air Stream (Parish et al., 2006) identified in the classification detailed in Coggins et al. (2014). Many of the patterns identified in the SOM are strongly influenced by the position of low-pressure systems and their corresponding cyclonic flows which is a well-known feature of the synoptic climatology of this region (Coggins & McDonald, 2015; Coggins et al., 2014). Work in Cohen et al. (2013) also shows that their Low Bellinghausen/Amundesen type was very common (30.8%) and was dominated by a low-pressure system in the Amundsen Sea area similar to our Node 12 in some respects. However, it is important to note that tests with larger and smaller number of nodes show that the key results presented later in this study are insensitive to choices about the classification.

The daily output from the eight different reanalyses is matched to the different nodes using a Euclidean distance metric based on a comparison of the winds from the current reanalyses relative to the anomaly

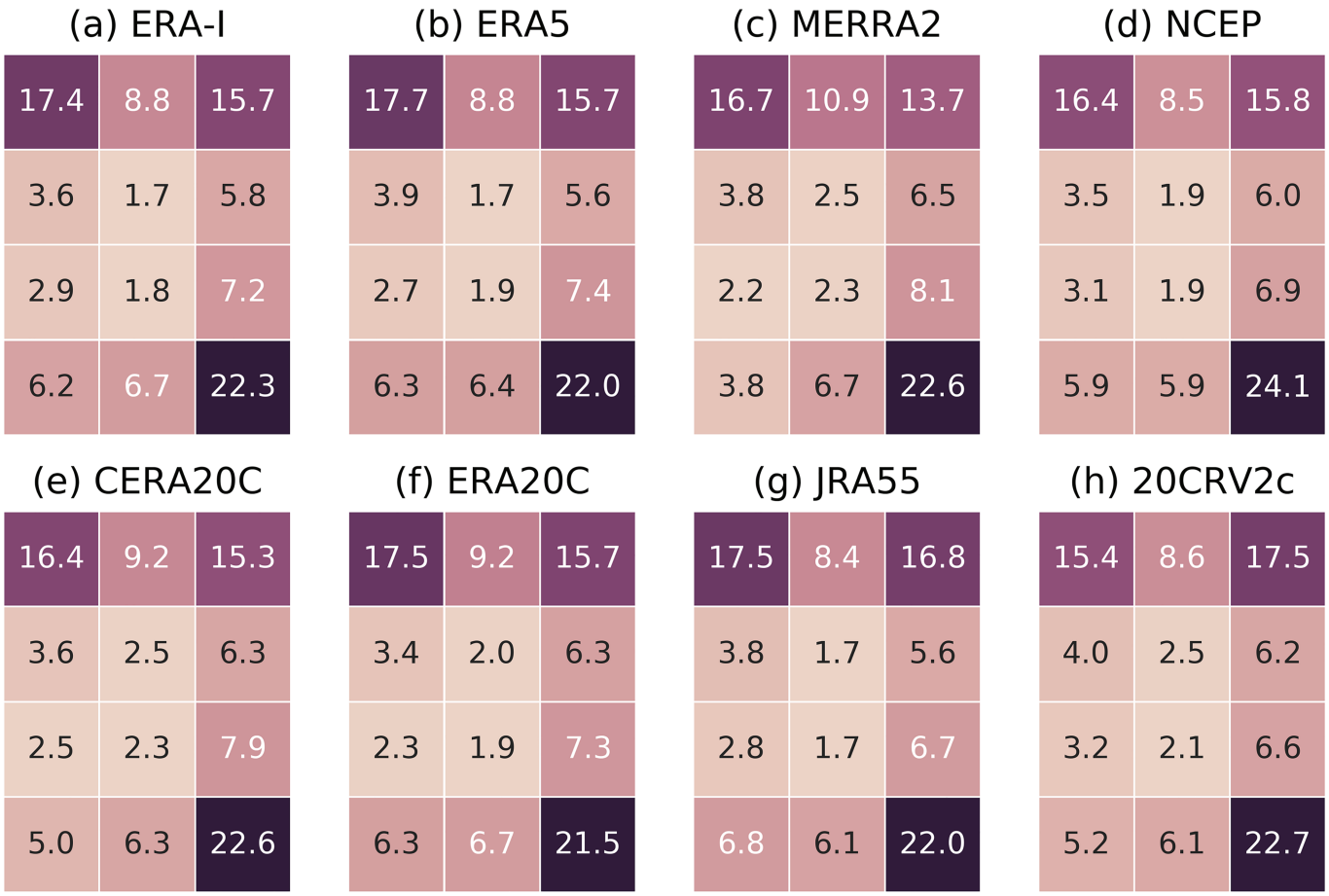

Figure 2. Relative frequency of occurrence for the 12 representative nodes for the eight reanalyses examined. Note that the ordering of nodes in each heatmap is the same as that identified in Figure 1. Note that modern reanalyses (ERA5, ERA-I, and MERRA2 have been grouped on the top row along with the NCEP-DOE reanalysis for reference purposes (a–d). Reanalyses products that provide output previous to 1979 (CERA20C, ERA20C, JRA55, and 20CRV2c) are displayed on the bottom row (e–h).

patterns output from the SOM applied to the ERA-Interim data added to the climatological mean pattern initially removed before the EOF process. The average relative frequency of occurrence (RFO) over the period 1980–1999 for each node for the eight different reanalyses are displayed in Figure 2. The frequency of occurrence of the patterns for these different reanalyses are rather similar. For example, the range for the most commonly occurring node (node 12 in every reanalysis) goes from 21.5% in ERA20C to 24.1% in NCEP–DOE (a 12% variation in relative terms). However, examination of the modern reanalyses (ERA5, ERA-Interim, MERRA2, and JRA55) displays a much smaller range of 22.0% to 22.6% (a 4% variation in relative terms) highlighting very good consistency. In addition, Nodes 1, 3, and 12 are the most commonly occurring nodes in every reanalysis, with only 20CRV2c having a higher RFO in Node 3 than in Node 1. There is more variability in relative terms in the less frequently occurring nodes, but again good consistency with Nodes 5 and 8 being the least frequently occurring nodes in all the reanalyses apart from MERRA2. For Node 8, RFO values vary between 1.7% and 2.5% (an approximately 32% change in relative terms). This result implies that in general the circulation patterns represented in the different reanalyses are rather similar and suggests that the reanalyses are very consistent over the period 1980–1999.

To further examine the consistency in the satellite era, we calculate the sum of the absolute differences in the RFO between each reanalysis, this acts as a simple measure of the consistency between reanalyses. Figure 3 displays a heatmap for all possible combinations of the eight reanalyses. Examination of Figure 3 shows that the ERA-Interim and the ERA5 output are most similar, the JRA55 and ERA20C reanalyses also show close correspondence with ERA-Interim and ERA5. Overall, ERA-Interim shows the smallest average absolute difference with all the other reanalyses, followed by ERA20C and then CERA20C. This is probably

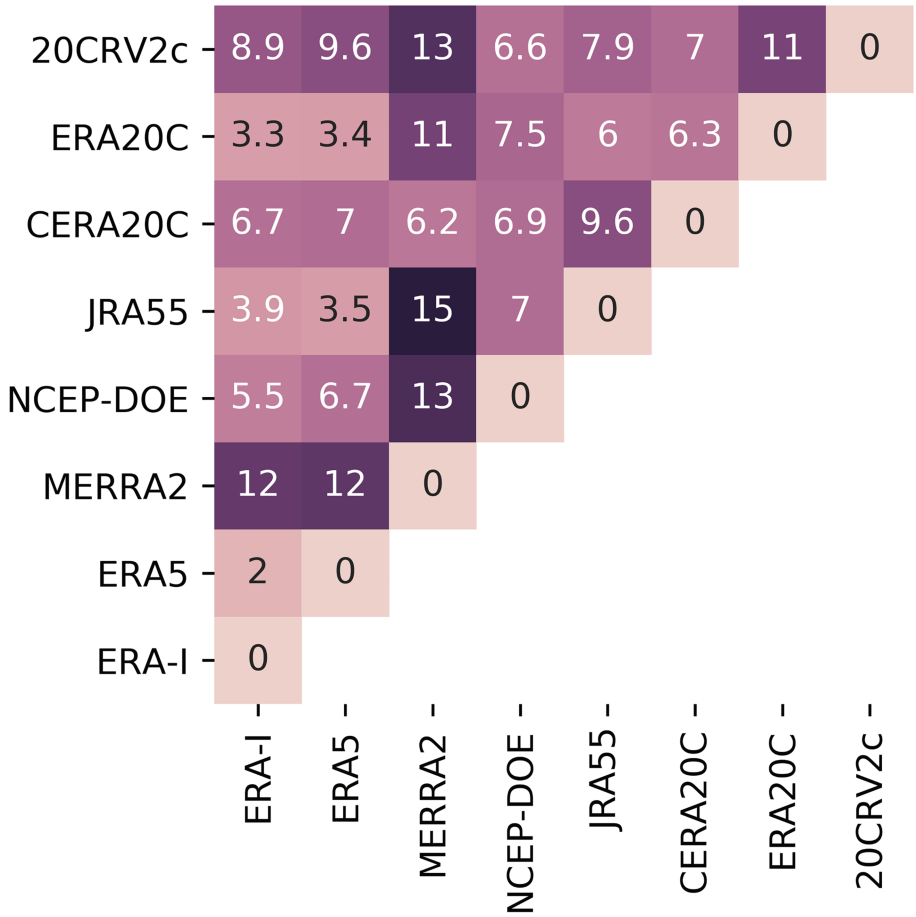

Figure 3. Sum of the absolute differences between each reanalyses relative frequency of occurrence from each node. Note that the diagonal line identifies zero difference because it relates to that reanalyses being compared with itself.

related to the similarities between the underlying models which all come from the ECMWF. Interestingly, the mostly poorly performing reanalysis using this simple metric is MERRA2, though the average of the sum of the absolute differences between MERRA2 and the other reanalyses is still only 8%.

This is an interesting result because a recent study shows that the quality of the MERRA2 reanalysis may be comparable to that of ERA-Interim in the Southern Hemisphere stratosphere (Friedrich et al., 2017). Figure 25 in Gelaro et al. (2017) also compares the MERRA and MERRA2 data sets with surface observations from Antarctic AWS and shows that the MERRA2 data set is an improvement over MERRA at high southern latitudes. So the MERRA2 product having the lowest overall consistency with ERA-Interim is somewhat surprising.

We now examine the year-to-year variation of the RFO for each node in each reanalysis to understand the interannual variability and also whether the RFO for particular nodes display any trends. The latter could represent a change in the synoptic climatology of the region, but a large number of studies have identified the potential for spurious trends in reanalyses in the polar regions (Befort et al., 2016; Bromwich & Fogt, 2004; Huai et al., 2019; Marshall et al., 2018; Schneider & Fogt, 2018; Screen & Simmonds, 2011; Wang et al., 2016; Wohland et al., 2019). Figure 4 displays the natural logarithm of the ratio of the RFO for a particular year divided by the mean over the 1980–1999 reference period. Note that each reanalysis is referenced to the RFO values displayed in Figure 2. Figure 4 displays these variations for the five full assimilation reanalyses (ERA-Interim, ERA5, JRA55, MERRA2, and NCEP-DOE) over the period 1980–2014. All of these reanalyses display similar time variations relative to the mean occurrence of that particular node with relatively small variations in the most commonly occurring nodes (Nodes 1, 3, and 12). The relative variations are larger in the less frequently occurring nodes, but even here there is relatively good agreement between the different

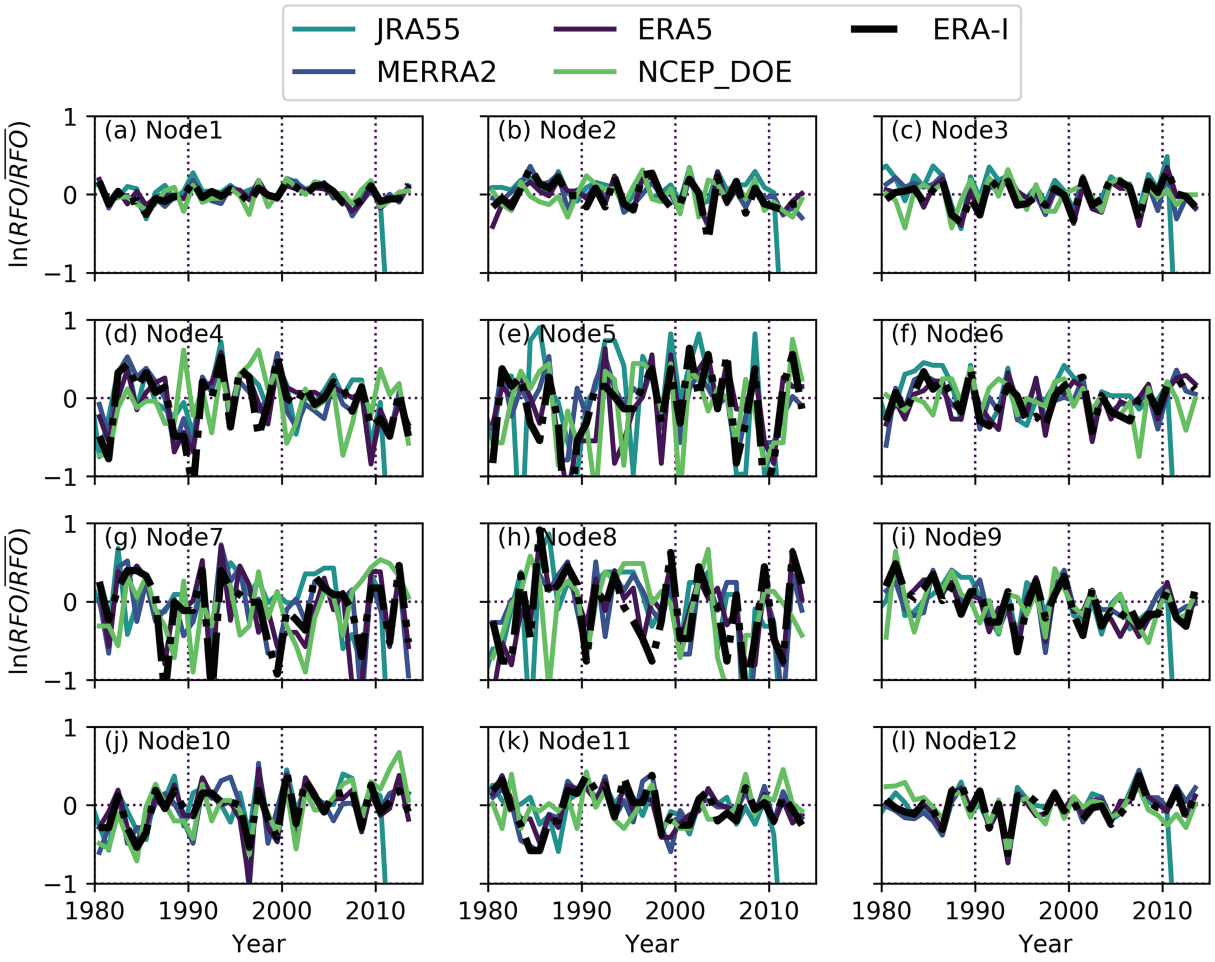

Figure 4. The natural logarithm of the ratio of the relative frequency of occurrence in a specific year divided by the 1980–1999 mean for each node for the five full assimilation reanalyses (ERA-Interim, ERA5, MERRA2, JRA55, and NCEP-DOE).

reanalyses. In summary, Figure 4 show good interannual correspondence between all the full assimilation reanalyses during the satellite era.

We now focus our attention on the reanalyses with output previous to the satellite era in Figure 5, the quality of each of these reanalyses is expected to reduce as we move to earlier periods because of the smaller quantities of observational data available. Examination of Figure 5 shows larger variability in these reanalyses. In particular, CERA20C and 20CRV2c show a breakpoint in the occurrence of Nodes 3, 7, and 10 at around 1957. In each of these three nodes the RFO is lower throughout the period previous to 1957 relative to the values in the satellite era. We also see marginally higher variability in the ERA20C data et before 1957. Given the dearth of observational data around the Antarctic previous to the International Geophysical Year in 1957 and previous studies we believe that this likely represents spurious trends in both these reanalyses which we will discuss later.

To further examine changes between various nodes, the root-mean-square difference (RMSD) between the ERA-Interim mean relative frequency of occurrence rates and the mean values for each reanalysis in each year are derived. Figure 6 displays the RMSD as a function of year for all eight reanalyses relative to the mean RFO values derived from ERA-Interim (shown in Figure 2). Figure 6 displays strong consistency up to 1979 in all the reanalyses supporting previous results detailed in Figures 2 and 3 with the average RMSD having a value of ~2%. For the four reanalyses with output before 1979 (CERA20C, ERA20C, JRA55, and 20CRV2c) there is a slightly larger RMSD between 1979 and 1957 of roughly 3%. For the three centennial-scale reanalyses, there is a distinct step-like increase previous to 1957 in the CERA20C (with a mean of 6%) and the 20CRV2c (with a mean of 5%) when the density of data assimilated in the region becomes very low. This

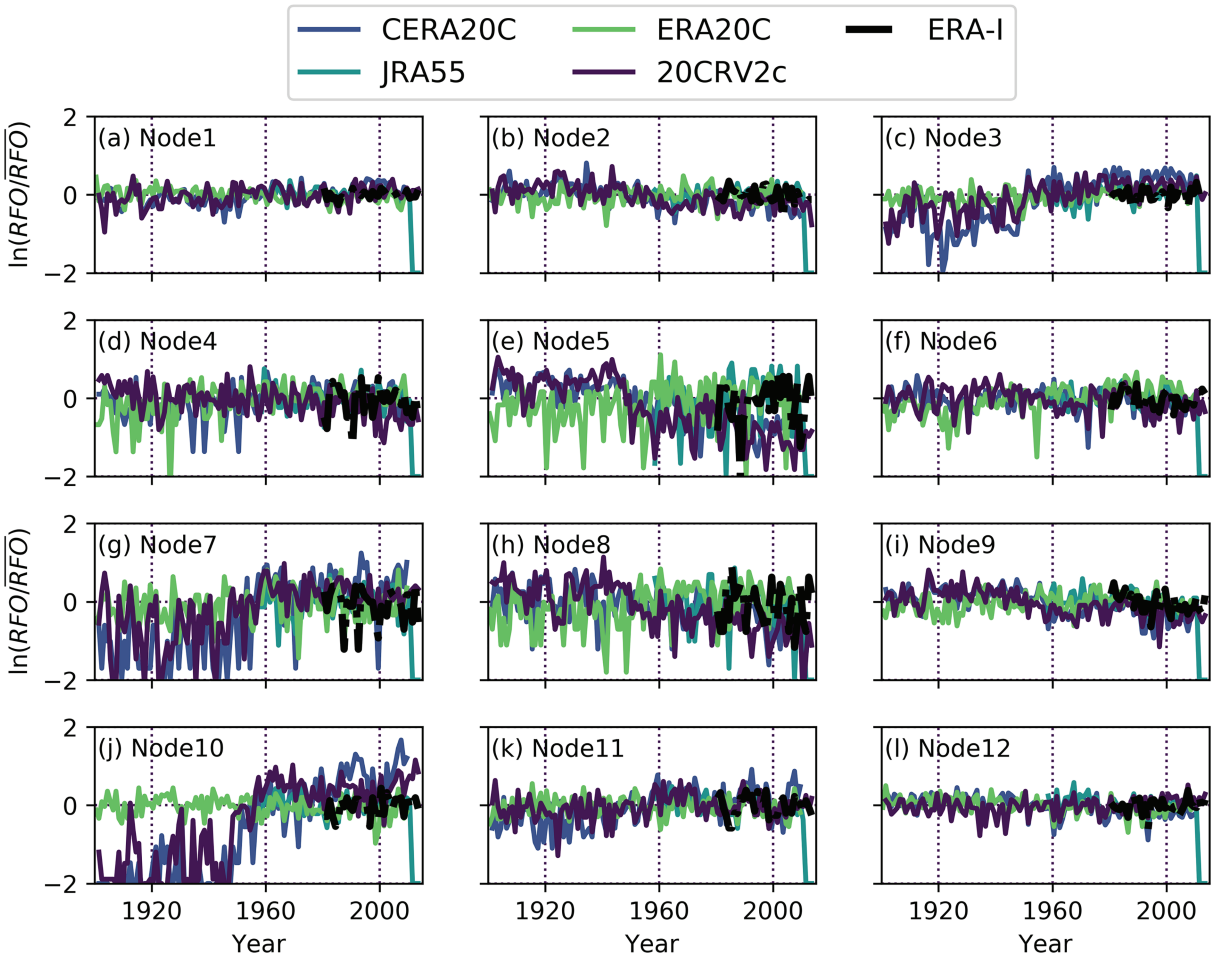

Figure 5. As in Figure 4 but for ERA-Interim, ERA20C, CERA20C, JRA55, and the 20CRV2c.

change clearly matches the patterns of change displayed in Figure 5. A more subtle increase in the ERA20C data set is also observed previous to 1957. The lack of consistency between these reanalyses suggests that these are likely to be spurious trends. Given previous studies, principally Schneider and Fogt (2018), we interpret these trends as associated with the relaxation of the reanalysis data back to the model climatology when observational constraints are removed.

Interestingly, Wohland et al. (2019) examined wind speed trends in four different centennial scale reanalyses globally. They concluded that assessments of long-term trends or climate variability from any single centennial reanalysis may be misleading. They suggest that the most accurate assessments would likely be based on an ensemble of centennial scale reanalyses and the results above are potentially supportive of this recommendation.

Thus far, the statistics examined have not considered whether the occurrence of a particular node occurs at the same time in different reanalyses. Figure 7 displays two example contingency tables which are used to examine the relationship between the categorical variables (the nodes) in each reanalyses. Figure 7 a displays the occurrence of a node in ERA-Interim contingent on the occurrence of nodes in ERA5. For Figure 7a we have calculated a normalized occurrence rate relative to the mean occurrence in ERA-Interim in percentage terms, this means that each column in the contingency table will sum to 100%. Note however that the rows do not have to sum to 100%. Inspection of Figure 7a shows a clear preference for the simultaneous occurrence of the same nodes in both ERA-Interim and ERA5 displayed by the diagonal nature of the matrix. Examination of the diagonal elements shows that the most frequently occurring nodes (Nodes 1, 3, and 12) also have the highest coincidence rates (near 90% in each case). This is not directly associated with their high frequency of occurrence as this factor has been removed in the normalization process, we suggest that this is likely

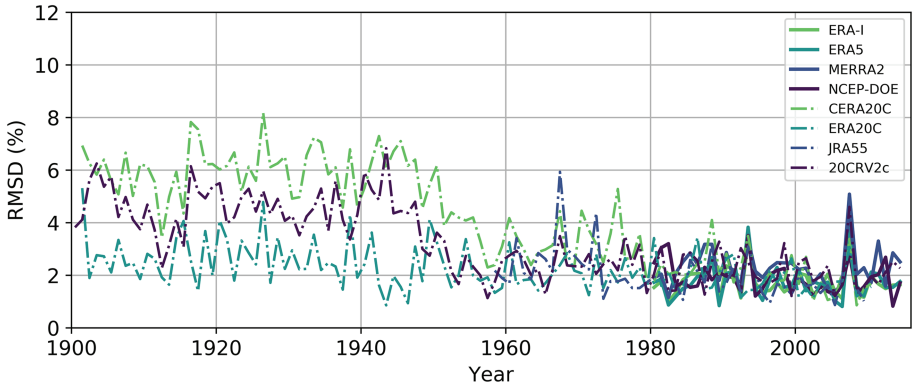

Figure 6. Root-mean-square difference between the yearly relative frequency of occurrence percentages for each reanalysis and the mean values identified in Figure 2 for ERA-Interim.

associated with the fact that these three nodes represent three very distinct synoptic states relative to their closest neighbors. Less frequently occurring nodes (Nodes 5, 7, and 8) display less coincidences, as identified in McDonald et al. (2016) these nodes may reflect transition states since the SOM classification scheme occasionally identifies transition nodes because the SOM creates a continuous gridded representation of the data space. We would therefore expect these transition states to span relatively empty portions of the physical data space, and because these states are temporary we might also expect lower coincidence.

Figure 7b displays the occurrence of a node in ERA-Interim contingent on the occurrence of nodes in 20CRV2c. Figure 7b still show some preference for the simultaneous occurrence of the same nodes in both reanalyses, but this is much weaker than in the contingency table displayed in Figure 7a. This is not surprising given the results displayed in Figures 2 and 3, but even the highest coincidence rates which are again connected to the most frequently occurring nodes (Nodes 1, 3, and 12) vary between 53.1% and 65.9%. Thus, while these two reanalyses display similar RFOs for the various nodes in their common period, their coincidence suggests much poorer consistency than at first sight. We also note that the 3 × 4 structure of the

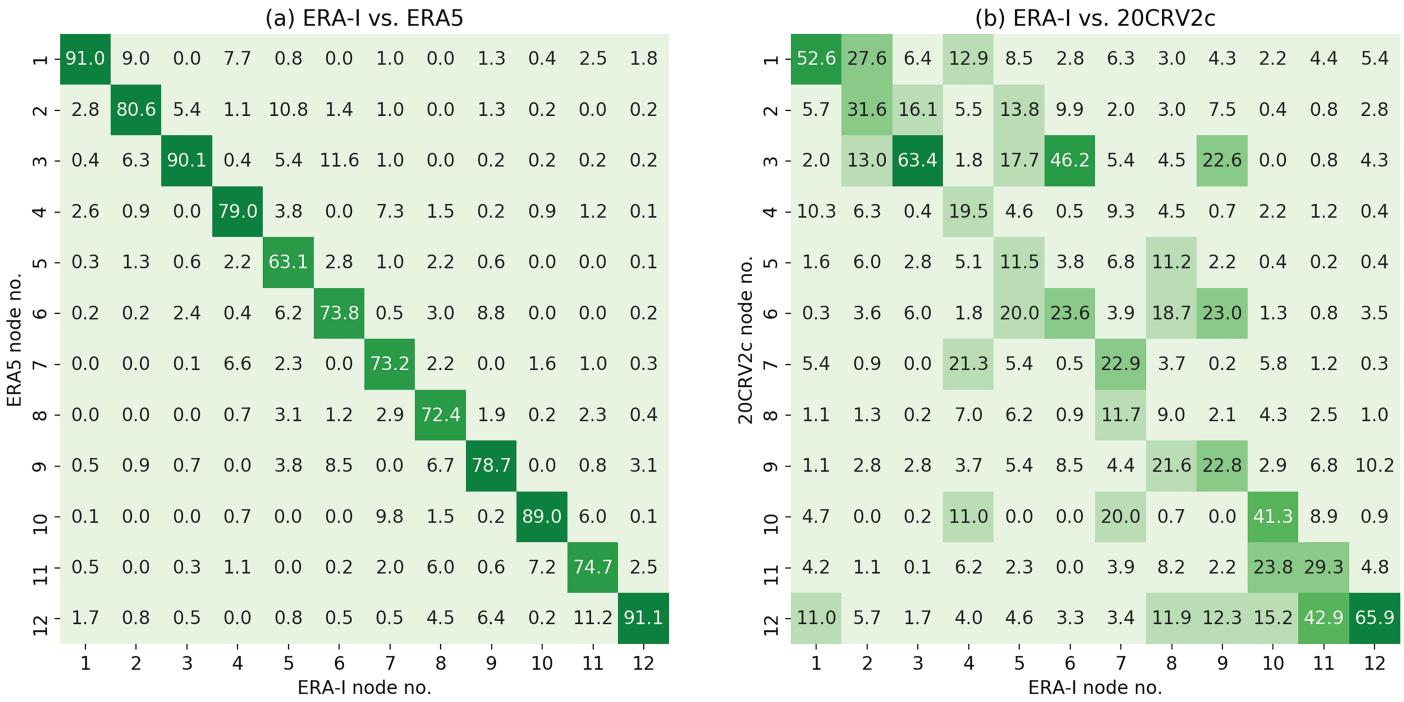

Figure 7. Contingency tables displaying the relative occurrence rate of particular nodes (in % terms) in the ERA-Interim and ERA5 reanalyses (a) and the ERA-Interim and the ERA20C reanalysis (b) occurring synchronously. The relative occurrence rate has been referenced to the ERA-Interim mean states, thus each column of the contingency table always adds to 100%.

SOM means that nodes in both nearby rows and columns are also observed in the contingency table. For example, Nodes 1 and 4 are close row neighbors and Nodes 1 and 2 would be close column neighbors. This means that secondary diagonals also occur, for example, Node 6 in ERA-Interim and Node 3 in the 20CRV2c occur frequently with a value of 46%, as does Node 9 in ERA-interim and Node 6 in 20CRV2c with a value of 23.0%, and these are both on the diagonal associated with nodes which are column neighbors.

The contingency tables displayed in Figure 7 are easy to interpret manually, but a single statistical metric to measure the overall association between the two categorical variables would be of value. McDonald and Parsons (2018) have previously used the entropy coefficient detailed in Press et al. (1992) to quantify the results from a contingency table. In this study, we expand on that effort, by deriving the entropy coefficient between each reanalyses as a function of time. The entropy coefficient defined in Press et al. (1992) varies between zero and one with zero representing no association and one representing perfect association between two classifications. A brief summary of the entropy coefficient is identified in Appendix A.

Figure 8 displays the entropy coefficient (U)as a function of time. Figure 8a displays the entropy coefficient for ERA-Interim relative to the seven other reanalyses over the period 1980–2014. Examination suggests that the entropy coefficient remains relatively constant or increases slightly as a function of time when ERA-Interim is compared to all the other reanalyses, with the entropy coefficient's value remaining relatively constant when ERA-Interim is compared to ERA5, MERRA2, NCEP-DOE, and 20CRV2c. The largest change in entropy coefficient over 1980–1999 occurs when ERA-Interim is compared with JRA55. Examination also suggests a wide range of consistency between the different reanalyses when compared to ERA-Interim. In particular, it is clear that the largest average U occurs between ERA-Interim and ERA5, but is closely followed by ERA-Interim and MERRA2. The centennial reanalyses clearly have lower associations with ERA-Interim than the current generation reanalyses. We also note that NCEP-DOE matches relatively poorly with ERA-Interim. It is also noteworthy that all the entropy coefficient values show a dip in 1998 which appears to be directly related to a large spurious outlier in the Ross Sea for a significant period in that year. We believe that this might represent a spurious buoy observation that has been assimilated by ERA-Interim in the RS which caused issues in June and July in our interpolated field.

We now examine the other reanalyses with short output periods, namely ERA5, MERRA2, and NCEP-DOE. Figure 8c displays the entropy coefficient for ERA5 relative to the seven other reanalyses over the period 1980–2014 and has very similar characteristics to the relationships displayed in Figure 8a. This is as might be expected given the extremely strong relationship between ERA5 and ERA-Interim as represented in Figures 3 and 7a. Figure 8e shows the entropy coefficient for MERRA2 relative to the seven other reanalyses over the period 1980–2014 and again has similar characteristics to the relationships displayed in Figures 8a and 8c. However, the average value of the entropy coefficient is slightly lower than those identified in Figure 8a and 8c. Inspection of the entropy coefficient in Figure 8g shows that the NCEP-DOE reanalyses shows much poorer consistency with the other reanalyses than the current generation reanalyses. Given that NCEP-DOE is an older generation with known issues, this is a strong demonstration of the entropy coefficient's value.

We now examine the reanalyses with longer output periods, namely CERA20C, ERA20C, JRA55, and 20CRV2c. The consistency between CERA20C and the seven other reanalyses over their common periods is displayed in Figure 8b. Examination of the entropy coefficient values shows that in the satellite era (1980–2014) there is considerable variance from year to year and between reanalyses compared with CERA20C, though the values vary between approximately 0.3 and 0.5 in the main. In the period between 1957 (the International Geophysical Year) and 1979 where a reasonable amount of data from the surface based observational network around Antarctica is available the values of the entropy coefficient also show significant variability. But, it is clear that for the three reanalyses that can be compared with CERA20C (ERA20C, JRA55, and 20CRV2c) the values of entropy coefficient are all lower than in the satellite era with a range of 0.1 to 0.3. Closer examination also shows that CERA20C matches best with the 20CRV2c reanalyses based on the entropy coefficient metric and the poorest association is between JRA55 and CERA20C in that period. Previous to 1957 the consistency between the three centennial reanalyses is very poor with entropy coefficient values below 0.2 in the main. Thus, Figure 8b can be characterized by increasing U as a function of time with very clear step changes around 1957 and 1979 which are highly likely to be associated with significant changes in the quantity of observations after the International Geophysical Year and at the start of the satellite era. This represents a very clear quantification of the quality of the reanalyses as a

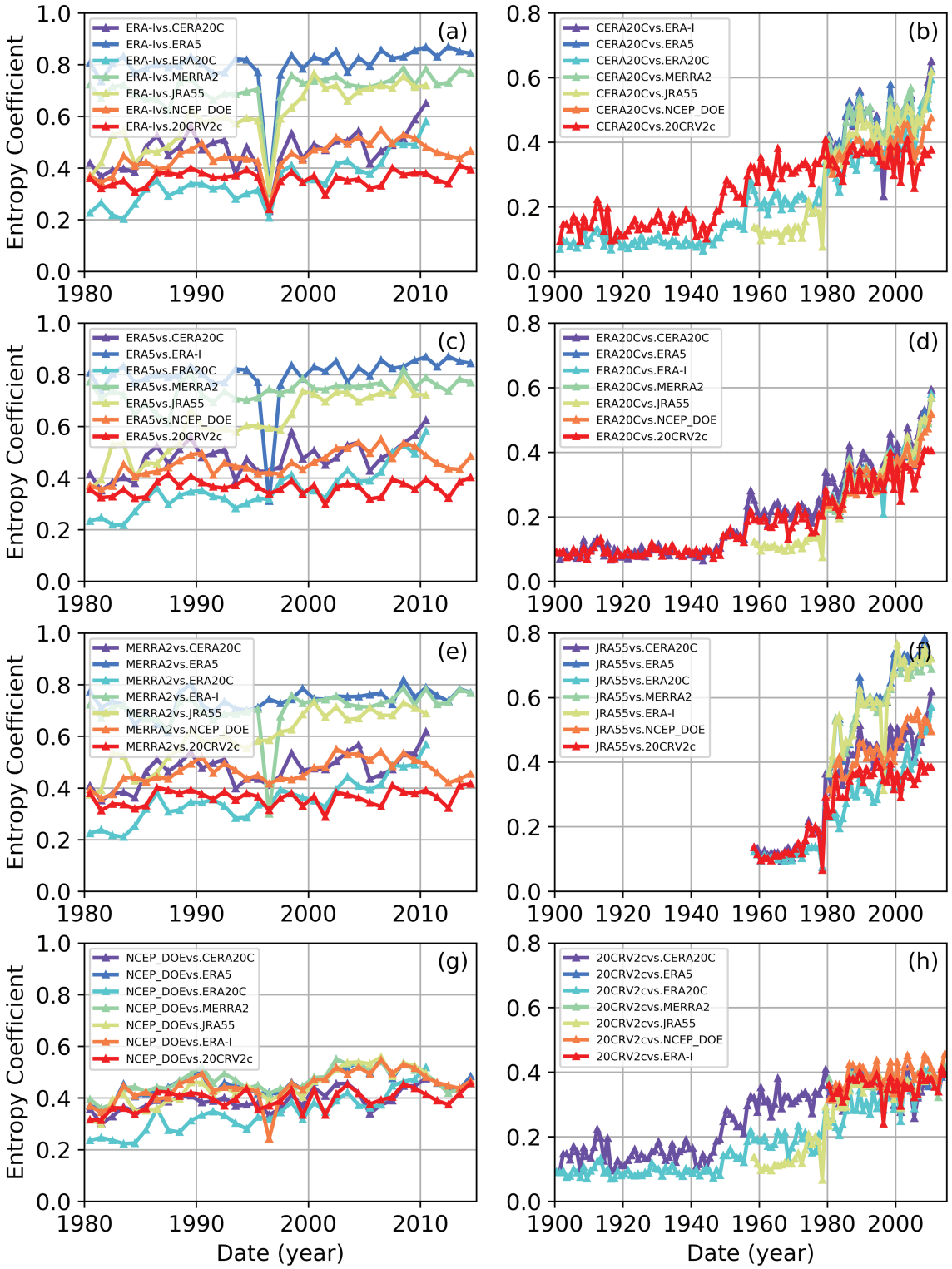

Figure 8. Time series of the entropy coefficient metric between different sets of reanalyses. For ERA-Interim (a), CERA20C (b), ERA5 (c), ERA20C (d), MERRA2 (e), JRA55 (f), NCEP-DOE (g), and 20CRV2c (h) against the seven other reanalyses. Note that the entropy coefficient lies between 0 and 1 with 0 indicating no association and 1 denoting a perfect association.

function of time in this region which shows that while the centennial reanalyses may be of value after 1957, they are highly inconsistent before the International Geophysical Year over this portion of the RIS/RS. It is interesting to note that Fogt et al. (2018) highlighted that compared to their statistical reconstruction of the pressure field south of 60° the twentieth century reanalyses (20CRV2c, ERA20C, and CERA20C) are similar for 20CR and ERA20C, but CERA20C agrees most closely with their reconstruction.

Figure 8d displays the entropy coefficient when ERA20C is compared to the seven other reanalyses over their common periods. Similarly to Figure 8b we see two-step changes around 1957 and 1979 which are almost certainly linked to changes in the quantity of observations assimilated after the International Geophysical Year and at the start of the satellite era. However, comparison of the entropy coefficient values in Figures 8b and 8d suggests that the CERA20C reanalysis matches more closely with other reanalyses than ERA20C. Figure 8f displays the entropy coefficient metric as a function of time for the JRA55 reanalysis. Note that the length of the record does not go all the way back to 1900 because the JRA55 record only covers the period from 1958 onward. Examination of Figure 8f shows a clear increase in the satellite era as we get nearer to the present which likely stems from increases in the quantity of satellite data assimilated into JRA55. This reanalysis displays a wide range of U in the satellite era with very high values when comparing with the other current generation reanalyses (ERA-Interim, ERA5, and MERRA2) and much lower values, representing lower consistency, with the centennial reanalyses and NCEP-DOE. Previous to the satellite era the entropy coefficients between MERRA2 and the three centennial reanalyses are all very low at around 0.1 and show almost no scatter. This suggests that the usage of the JRA55 reanalysis previous to the satellite era is unlikely to be a good option.

Figure 8h displays the entropy coefficient for the 20CRV2c reanalysis relative to the seven other reanalyses over the period 1900–2014. In this case there are again signs of the two steps in the values of the entropy coefficient around the International Geophysical Year and in 1979. Notably, 20CRV2c generally matches more closely to CERA20C than ERA20C previous to the satellite era. Given the results in Figure 3 which suggest that the series of reanalysis from ECMWF match each other most closely this is a little unexpected. However, we suggest that this relationship is associated with the changes in RFO for different nodes in Figure 5 changing in similar ways previous to the International Geophysical Year for 20CRV2c and ERA20C.

To highlight the usefulness of the entropy coefficient we compare this metric with the ensemble spread from the 20CRV2c and CERA20C in Figure 9. In this analysis we use the ensemble standard deviation applied to wind speed data as defined in Grimit and Mass (2007) as our measure of the ensemble spread. Figure 9a shows the entropy coefficient for CERA20C against other centennial renalyses (ERA20C and 20CRV2c) and the corresponding ensemble spread. Note that we display the average ensemble spread for a particular year over the region identified in Figure 1. Examination of the CERA20C ensemble spread derived from the 56 ensemble members has large near-constant values between 1901 and around 1957, slightly lower values between 1957 and 1979 and values that decrease monotonically after 1979. These changes are very obviously related to the quantity of data assimilated into that reanalyses. Comparison with the two entropy coefficient calculations displays a near-opposite response. Thus, there is a clear negative relationship between the entropy coefficient and the ensemble spread in this case. Though, analysis (not shown) indicates that this is not a simple linear relationship. Figure 9b shows the entropy coefficient for 20CRV2c against other centennial reanalyses (ERA20C and CERA20c) and the corresponding ensemble spread. Again a clear negative relationship is observed, though it is clear that this is again not a simple linear relationship. Given that the ensemble spread is a measure of the self-consistency of a particular reanalyses and the entropy coefficient is a measure of the consistency between reanalyses it is not surprising that the relationship is not linear. The complex relationship between the two measures may also be related to the fact that the ensemble spread does not measure the model uncertainty component, namely the uncertainty associated with the underlying assumptions within the model, which is likely largest in regions with few or no observations. However, Figure 9 clearly demonstrates that the entropy coefficient can be used to identify the uncertainty associated with a reanalyses relative to a current well-used measure. It also demonstrates that the entropy coefficient is a useful measure of the uncertainty associated with a specific reanalyses. Interestingly, the ensemble spread in the CERA20C is nearly a factor of 2 larger than that in the 20CRV2c reanalyses for the period between 1901 and 1957. This large difference is not reflected in the corresponding entropy coefficients.

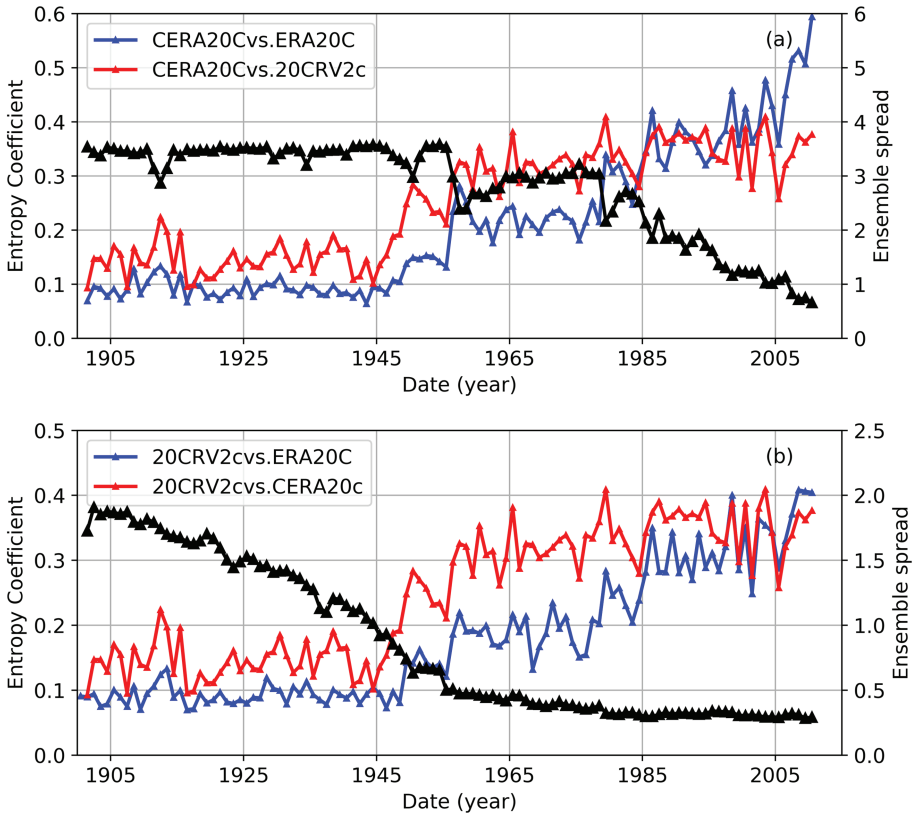

Figure 9. Time series of the entropy coefficient metric between different sets of centennial reanalyses as identified in the legend and corresponding ensemble spread information (black line) are displayed. Entropy coefficients and ensemble spreads for the CERA20C (a) and 20CRV2c (b) are displayed.

4. Discussion and Conclusions

This study compares the consistency in the surface winds in eight reanalyses, namely, CERA20C, ERA20C, ERA5, ERA-Interim, JRA55, NCEP-DOE, MERRA2, and 20CRV2c. Development of a SOM classification of surface winds using ERA-Interim output over the common period (1980–1999) was used to define representative synoptic weather states (identified as nodes in the SOM) over the RS/RIS. Though, we note that very similar patterns resulted when we used other reanalyses for our reference. The mean frequency of occurrence of these synoptic patterns for the CERA20C and 20CRV2c reanalyses changes previous to 1957, where clear decreasing trends in some nodes are observed. Given the large amount of literature that has identified spurious trends previous to the satellite era in reanalyses, it appears likely that these trends are associated with the models relaxing to their climatology because of a lack of observational constraints. For example, Schneider and Fogt (2018) recently compared three centennial reanalyses around Antarctica previous to the satellite era and found that the quality of the pressure field is compromised by the lack of observations relative to a pressure field reconstruction. Huai et al. (2019) also compared a number of reanalyses with temperature observations over Antarctica and showed that all the centennial-scale reanalyses show trends previous to 1979 which are likely to be spurious. The work by Huai et al. (2019) which examines a number of reanalyses over Antarctica also identifies shifts in 1979 similar to our Figure 8. We speculate that this shift might be associated with data sets that have been used in the reanalyses other than the surface pressure observations, such as the HadISST data set (Rayner et al., 2003), which shows a sudden decreases in the sea ice concentration in the late 1970s. This idea has previously been identified in Huai et al. (2019). Figure 9 in the current study also shows a clear pattern of increasing ensemble spread in the CERA20C up to 1979 and then a clear step shift at around 1957. Our conclusion is also supported by studies which showed that the representation of extratropical cyclones in the 20CRV2 and ERA20C reanalysis was substantially different between the two data sets, especially in the first half of the twentieth century (Befort et al., 2016).

Most importantly, this work introduces the entropy coefficient metric as a way to quantify the consistency in the representation of particular synoptic states between reanalyses. Noting that the entropy coefficient varies between zero and one with zero representing no association and one representing perfect association between two reanalyses. Examination of Figure 8 identifies very clearly that all the current generation reanalyses (ERA5, ERA-Interim, JRA55, and MERRA2) are highly consistent in the satellite era, though the performance of JRA55 reduces as we move further back in the satellite era. Given other results we therefore recommend the usage of ERA-Interim and ERA5 over this region in the satellite era and suggest some caution using MERRA2, which has a less consistent climatological state than that represented in the other reanalyses. Though, we identify this conclusion cautiously as the entropy coefficient is a metric of the consistency between data sets and in the absence of significant independent observational data to define the reference standard, it may be that MERRA2 is actually the most representative of the modern reanalyses.

Inspection of Figures 8b, 8d, 8f, and 8h suggests that all four of the reanalyses that provide output previous to the satellite era (CERA20C, ERA20C, JRA55, and 20CRV2c) display a step change in consistency as measured by the entropy coefficient at around 1979. Though, it is notable that the magnitude of this step is quite small when CERA20C and 20CRV2c are compared. In addition, all of the centennial reanalyses (CERA20C, ERA20C, and 20CRV2c) show a very significant step change between periods before and after the International Geophysical Year (1957). Given previous work and general discussion as detailed in Thorne and Vose (2010) we believe that this is a clear sign of the influence of the significantly reduced density of observations around Antarctica previous to 1957. Given the very low values of the entropy coefficient before 1957, we conclude that none of the centennial reanalyses can be trusted in this region of Antarctica before the International Geophysical Year.

Figure 9 compares the entropy coefficient values against measures of the ensemble spread for the CERA20C and 20CRV2c reanalyses. A very clear negative, but not necessarily linear, relationship between the entropy coefficient and the ensemble spreads are observed. This demonstrates that the entropy coefficient is a useful measure of the uncertainty associated with a specific reanalyses. The increase in the ensemble spread as data density decreases in this region also supports our conclusion that the centennial reanalyses are of limited utility in this region of Antarctica before the International Geophysical Year.

We also note that for studies focused on trends the centennial reanalyses might be useful over the period 1957 to 2014; however, they all display relatively low consistency with the current generation of reanalyses in the satellite era. We therefore suggest that the centennial reanalyses should be treated cautiously when examining climate variability in the satellite era in data sparse regions in general. Given that reanalyses are widely applied in the polar regions to study climate variability and trends, to understand large-scale circulation patterns, to validate climate models, and to provide boundary conditions for ocean/sea ice models this is a significant issue. This study therefore highlights the importance of data rescue efforts, such as the Atmospheric Circulation Reconstruction over the Earth (ACRE) Initiative (Allan et al., 2011), and the continued usage of paleoclimate proxies in helping to understand the change in this region as identified previously in Jones et al. (2016).

More generally, this study demonstrates the ability of the entropy coefficient to act as a proxy for the quality of reanalyses. However, we would suggest that the calculation of the entropy coefficient is most useful when applied to multiple reanalyses, this aligns in some ways with the suggestion in Wohland et al. (2019) that the most accurate assessments of trends would likely be based on an ensemble of centennial-scale reanalyses. This methodology is also likely to be most useful in combination with more traditional studies which compare observational records that have not been assimilated into the reanalyses examined. However, the limited quantities of independent observations available in the polar regions (Jakobson et al., 2012; Nygard et al., 2016) may make this method particularly valuable in these regions and over other data sparse regions such as the oceans.

Additionally, previous work by Torralba et al. (2017) which focused on an assessment of the ability of atmospheric reanalyses to reproduce wind speed trends has identified that to reduce the uncertainty in long-term trends the use of more than one reanalysis is beneficial and the entropy coefficient measure may provide a way to determine whether certain reanalyses should be excluded from that ensemble in specific areas.

Work to further highlight the utility of the entropy coefficient will focus on the application of this analysis to other locations around the Antarctic continent and globally. There is also the potential to expand this

methodology by considering the multiple SOM classification approach used in Gibson et al. (2016). The potential to directly link the quantity of historical observations to the quality of the reanalyses output is also an obvious future avenue. Another area of future work would be to redo this analysis with new reanalysis products, for example, the ERA5 reanalysis data are currently only available back to 1979, but are planned to go back to 1950. A third version of the 20CR which will replace 20CRV2c is also currently in production.

Appendix A: Entropy Coefficient

This section provides a short summary of the entropy coefficient metric used in this work. A more detailed description is available in Theil (1970) and Press et al. (1992). The entropy coefficient, also called the uncertainty coefficient, proficiency, or Theil's U is a measure of the association between two categorical variables. In this study the synoptic state observed in two reanalyses which could be referred to as x and y, is measured. For values between 0 and 1, the entropy coefficient gives the fraction of variable Y's entropy that is lost if variable X is already known. In the current case the entropy coefficient effectively quantifies the frequency that two reanalyses identify the same node at the same time relative to the value expected by random chance.

The entropy coefficient can be calculated when we have samples from two discrete random variables X and Y. If variable X has I possible values and Y has J then the entropy of variable X can be written as

$$H(X) = -\sum_{i=1}^{I} p_i \ln(p_i) \tag{A1}$$

where p_i is equal to N_i/N where N is the total number of observations and N_i is the number of observations which are categorized into state i. The entropy is one way to summarize the structure in the probability density function and is a measure of information content of the variables.

To derive the entropy coefficient, we also use the conditional entropy ($H(Y|X)$) which quantifies the amount of information needed to describe variable Y given that the value of variable X is known:

$$H(Y|X) = -\sum_i p_i \sum_j \frac{p_{ij}}{p_i} \ln \frac{p_{ij}}{p_i} = -\sum_{ij} p_{ij} \ln \frac{p_{ij}}{p_i} \tag{A2}$$

where p_{ij} is equal to N_{ij}/N where N_{ij} is the number of observations which are categorized into states i and j.

We can now define the entropy coefficient of Y on X. This lies between 0 and 1 with 0 indicating no association and 1 denoting complete association. For any values in between these extreme, $U(Y|X)$ gives the proportion of Y's entropy that is lost if X is already known, where $U(Y|X)$ is defined by

$$U(Y|X) = \frac{H(Y) - H(Y|X)}{H(Y)} \tag{A3}$$

The entropy coefficient is a measure of comparison, but unfortunately is not symmetric. We therefore use the symmetric version of the entropy coefficient, which is the weighted average of $U(Y|X)$ and $U(X|Y)$ and can be written as

$$U(X,Y) = 2\left[\frac{H(Y) + H(X) - H(X,Y)}{H(X) + H(Y)}\right] \tag{A4}$$

If both X and Y are independent, then $H(X,Y) = H(X) + H(Y)$ and equation (A4) is zero. In the case of complete dependence, $H(X,Y) = H(X) = H(Y)$ and equation (A4) equals one. We should also identify that the entropy coefficient has been used previously for measuring the validity of statistical classification algorithms.

Acknowledgments
The author would like to acknowledge the support associated with the NZARI RFP 2014-2 Vulnerability of the Ross Ice Shelf in a Warming World project and internal funds from the University of Canterbury. We would also like to acknowledge The JRA55 data set used for this study is provided from the Japanese 55-year Reanalysis project carried out by the Japan Meteorological Agency (JMA). We would also like to acknowledge the ERA–Interim, ERA5, ERA20C and CERA20C data sets provided by the European Centre for Medium Range Weather Forecasting. The NCEP-DOE and 20CRV2c reanalysis data sets are provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at https://www.esrl.noaa.gov/psd/. We would also like to acknowledge MDISC, managed by the NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC), who provide access to the MERRA2 data set.

References

Allan, R., Brohan, P., Compo, G. P., Stone, R., Luterbacher, J., & Brönnimann, S. (2011). The International Atmospheric Circulation Reconstructions over the Earth (ACRE) initiative. *Bulletin of the American Meteorological Society*, *92*(11), 1421–1425. https://doi.org/10.1175/2011bams3218.1

Barthelemy, A., Goosse, H., Fichefet, T., & Lecomte, O. (2018). On the sensitivity of Antarctic sea ice model biases to atmospheric forcing uncertainties. *Climate Dynamics*, *51*(4), 1585–1603. https://doi.org/10.1007/s00382-017-3972-7

Befort, D. J., Wild, S., Kruschke, T., Ulbrich, U., & Leckebusch, G. C. (2016). Different long-term trends of extra-tropical cyclones and windstorms in ERA-20C and NOAA-20CR reanalyses. *Atmospheric Science Letters*, *17*(11), 586–595. https://doi.org/10.1002/asl.694

Bracegirdle, T. J. (2013). Climatology and recent increase of westerly winds over the Amundsen Sea derived from six reanalyses. *International Journal of Climatology*, *33*(4), 843–851. https://doi.org/10.1002/joc.3473

Bracegirdle, T. J., & Marshall, G. J. (2012). The reliability of Antarctic tropospheric pressure and temperature in the latest global reanalyses. *Journal of Climate*, *25*(20), 7138–7146. https://doi.org/10.1175/jcli-d-11-00685.1

Bromwich, D. H., & Fogt, R. L. (2004). Strong trends in the skill of the ERA-40 and NCEP-NCAR reanalyses in the high and midlatitudes of the Southern Hemisphere, 1958–2001. *Journal of Climate*, *17*(23), 4603–4619. https://doi.org/10.1175/3241.1

Bromwich, D. H., Fogt, R. L., Hodges, K. I., & Walsh, J. E. (2007). A tropospheric assessment of the ERA-40, NCEP, and JRA-25 global reanalyses in the polar regions. *Journal of Geophysical Research*, *112*, D10111. https://doi.org/10.1029/2006jd007859

Bromwich, D. H., Nicolas, J. P., & Monaghan, A. J. (2011). An assessment of precipitation changes over Antarctica and the Southern Ocean since 1989 in contemporary global reanalyses. *Journal of Climate*, *24*(16), 4189–4209. https://doi.org/10.1175/2011jcli4074.1

Coggins, J. H. J., & McDonald, A. J. (2015). The influence of the Amundsen Sea Low on the winds in the Ross Sea and surroundings: Insights from a synoptic climatology. *Journal of Geophysical Research: Atmospheres*, *120*, 2167–2189. https://doi.org/10.1002/2014jd022830

Coggins, J. H. J., McDonald, A. J., & Jolly, B. (2014). Synoptic climatology of the Ross Ice Shelf and Ross Sea region of Antarctica: *k*-means clustering and validation. *International Journal of Climatology*, *34*(7), 2330–2348. https://doi.org/10.1002/joc.3842

Cohen, L., Dean, S., & Renwick, J. (2013). Synoptic weather types for the Ross Sea region, Antarctica. *Journal of Climate*, *26*(2), 636–649. https://doi.org/10.1175/jcli-d-11-00690.1

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., et al. (2011). The twentieth century reanalysis project. *Quarterly Journal of the Royal Meteorological Society*, *137*(654), 1–28.

Dale, E. R., McDonald, A. J., Coggins, J. H. J., & Rack, W. (2017). Atmospheric forcing of sea ice anomalies in the Ross Sea polynya region. *Cryosphere*, *11*(1), 266–280. https://doi.org/10.5194/tc-11-267-2017

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828

Fogt, R. L., Clark, L. N., & Nicolas, J. P. (2018). A new monthly pressure dataset poleward of 60s since 1957. *Journal of Climate*, *31*(10), 3865–3874. https://doi.org/10.1175/jcli-d-17-0879.1

Fogt, R. L., Goergens, C. A., Jones, J. M., Schneider, D. P., Nicolas, J. P., Bromwich, D. H., & Dusselier, H. E. (2017). A twentieth century perspective on summer Antarctic pressure change and variability and contributions from tropical SSTs and ozone depletion. *Geophysical Research Letters*, *44*, 9918–9927. https://doi.org/10.1002/2017gl075079

Fogt, R. L., Goergens, C. A., Jones, M. E., Witte, G. A., Lee, M. Y., & Jones, J. M. (2016). Antarctic station-based seasonal pressure reconstructions since 1905: 1. Reconstruction evaluation. *Journal of Geophysical Research: Atmospheres*, *121*, 2814–2835. https://doi.org/10.1002/2015jd024564

Friedrich, L. S., McDonald, A. J., Bodeker, G. E., Cooper, K. E., Lewis, J., & Paterson, A. J. (2017). A comparison of loon balloon observations and stratospheric reanalysis products. *Atmospheric Chemistry and Physics*, *17*(2), 855–866. https://doi.org/10.5194/acp-17-855-2017

Gelaro, R., McCarty, W., Suarez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. https://doi.org/10.1175/jcli-d-16-0758.1

Gibson, P. B., Uotila, P., Perkins-Kirkpatrick, S. E., Alexander, L. V., & Pitman, A. J. (2016). Evaluating synoptic systems in the CMIP5 climate models over the Australian region. *Climate Dynamics*, *47*(7-8), 2235–2251. https://doi.org/10.1007/s00382-015-2961-y

Grimit, E. P., & Mass, C. F. (2007). Measuring the ensemble spread-error relationship with a probabilistic approach: Stochastic ensemble results. *Monthly Weather Review*, *135*(1), 203–221. https://doi.org/10.1175/mwr3262.1

Hosking, J. S., Orr, A., Marshall, G. J., Turner, J., & Phillips, T. (2013). The influence of the Amundsen-Bellingshausen seas low on the climate of West Antarctica and its representation in coupled climate model simulations. *Journal of Climate*, *26*(17), 6633–6648. https://doi.org/10.1175/jcli-d-12-00813.1

Hoskins, B. J., & Hodges, K. I. (2005). A new perspective on Southern Hemisphere storm tracks. *Journal of Climate*, *18*(20), 4108–4129. https://doi.org/10.1175/jcli3570.1

Huai, B., Wang, Y., Ding, M., Zhang, J., & Dong, X. (2019). An assessment of recent global atmospheric reanalyses for Antarctic near surface air temperature. *Atmospheric Research*, *226*, 181–191. https://doi.org/10.1016/j.atmosres.2019.04.029

Jakobson, E., Vihma, T., Palo, T., Jakobson, L., Keernik, H., & Jaagus, J. (2012). Validation of atmospheric reanalyses over the central Arctic Ocean. *Geophysical Research Letters*, *39*, L10802. https://doi.org/10.1029/2012gl051591

Jolly, B., Kuma, P., McDonald, A., & Parsons, S. (2018). An analysis of the cloud environment over the Ross Sea and Ross Ice Shelf using CloudSat/CALIPSO satellite observations: The importance of synoptic forcing. *Atmospheric Chemistry and Physics*, *18*(13), 9723–9739. https://doi.org/10.5194/acp-18-9723-2018

Jolly, B., McDonald, A. J., Coggins, J. H. J., Zawar-Reza, P., Cassano, J., Lazzara, M., et al. (2016). A validation of the Antarctic mesoscale prediction system using self-organizing maps and high-density observations from snowweb. *Monthly Weather Review*, *144*(9), 3181–3200. https://doi.org/10.1175/mwr-d-15-0447.1

Jones, J. M., Gille, S. T., Goosse, H., Abram, N. J., Canziani, P. O., Charman, D. J., et al. (2016). Assessing recent trends in high-latitude Southern Hemisphere surface climate. *Nature Climate Change*, *6*(10), 917–926. https://doi.org/10.1038/nclimate3103

Jones, R. W., Renfrew, I. A., Orr, A., Webber, B. G. M., Holland, D. M., & Lazzara, M. A. (2016). Evaluation of four global reanalysis products using in situ observations in the Amundsen Sea Embayment, Antarctica. *Journal of Geophysical Research: Atmospheres*, *121*, 6240–6257. https://doi.org/10.1002/2015jd024680

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S. K., Hnilo, J. J., Fiorino, M., & Potter, G. L. (2002). NCEP-DOE AMIP-II reanalysis (R-2). *Bulletin of the American Meteorological Society*, *83*(11), 1631–1643. https://doi.org/10.1175/bams-83-11-1631

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., et al. (2015). The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan*, *93*(1), 5–48. https://doi.org/10.2151/jmsj.2015-001

Kohonen, T. (1990). The self-organizing map. *Proceedings of the Ieee*, *78*(9), 1464–1480. https://doi.org/10.1109/5.58325

Laloyaux, P., de Boisseson, E., Balmaseda, M., Bidlot, J. R., Broennimann, S., Buizza, R., et al. (2018). CERA-20C: A coupled reanalysis of the twentieth century. *Journal of Advances in Modeling Earth Systems*, *10*, 1172–1195. https://doi.org/10.1029/2018ms001273

Lazzara, M. A., Weidner, G. A., Keller, L. M., Thom, J. E., & Cassano, J. J. (2012). Antarctic automatic weather station program 30 years of polar observations. *Bulletin of the American Meteorological Society*, *93*(10), 1519–1537. https://doi.org/10.1175/bams-d-11-00015.1

Li, M., Liu, J. P., Wang, Z. Z., Wang, H., Zhang, Z. H., Zhang, L., & Yang, Q. H. (2013). Assessment of sea surface wind from NWP reanalyses and satellites in the Southern Ocean. *Journal of Atmospheric and Oceanic Technology*, *30*(8), 1842–1853. https://doi.org/10.1175/jtech-d-12-00240.1

Lindsay, R., Wensnahan, M., Schweiger, A., & Zhang, J. (2014). Evaluation of seven different atmospheric reanalysis products in the Arctic*. *Journal of Climate*, *27*(7), 2588–2606. https://doi.org/10.1175/jcli-d-13-00014.1

Liu, Y. H., & Key, J. R. (2016). Assessment of Arctic cloud cover anomalies in atmospheric reanalysis products using satellite data. *Journal of Climate*, *29*(17), 6065–6083. https://doi.org/10.1175/jcli-d-15-0861.1

Marshall, G. J., Kivinen, S., Jylha, K., Vignols, R. M., & Rees, W. G. (2018). The accuracy of climate variability and trends across Arctic Fennoscandia in four reanalyses. *International Journal of Climatology*, *38*(10), 3878–3895. https://doi.org/10.1002/joc.5541

McDonald, A. J., Cassano, J. J., Jolly, B., Parsons, S., & Schuddeboom, A. (2016). An automated satellite cloud classification scheme using self-organizing maps: Alternative ISCCP weather states. *Journal of Geophysical Research: Atmospheres*, *121*, 13,009–13,030. https://doi.org/10.1002/2016jd025199

McDonald, A. J., & Parsons, S. (2018). A comparison of cloud classification methodologies: Differences between cloud and dynamical regimes. *Journal of Geophysical Research: Atmospheres*, *123*, 11,173–11,193. https://doi.org/10.1029/2018jd028595

Nicolas, J. P., & Bromwich, D. H. (2011). Precipitation changes in high southern latitudes from global reanalyses: A cautionary tale. *Surveys in Geophysics*, *32*, 475–494. https://doi.org/10.1007/s10712-011-9114-6

Nigro, M. A., & Cassano, J. J. (2014). Analysis of the Ross Ice Shelf airstream forcing mechanisms using self-organizing maps. *Monthly Weather Review*, *142*(12), 4719–4734. https://doi.org/10.1175/mwr-d-14-00077.1

Nygard, T., Vihma, T., Birnbaum, G., Hartmann, J., King, J., Lachlan-Cope, T., et al. (2016). Validation of eight atmospheric reanalyses in the Antarctic Peninsula region. *Quarterly Journal of the Royal Meteorological Society*, *142*(695), 684–692. https://doi.org/10.1002/qj.2691

Parish, T. R., Cassano, J. J., & Seefeldt, M. W. (2006). Characteristics of the Ross Ice Shelf air stream as depicted in Antarctic mesoscale prediction system simulations. *Journal of Geophysical Research*, *111*, D12109. https://doi.org/10.1029/2005jd006185

Parsons, S., Renwick, J. A., & McDonald, A. J. (2016). An assessment of future Southern Hemisphere blocking using CMIP5 projections from four GCMs. *Journal of Climate*, *29*(21), 7599–7611. https://doi.org/10.1175/jcli-d-15-0754.1

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., et al. (2016). ERA-20C: An atmospheric reanalysis of the twentieth century. *Journal of Climate*, *29*(11), 4083–4097. https://doi.org/10.1175/jcli-d-15-0556.1

Press, W. H., Teukolsky, S. A., Vettering, W. T., & Flannery, B. P. (1992). *Numerical recipes in Fortran 77* (2nd ed.). Cambridge: Cambridge University Press.

Rapaic, M., Brown, R., Markovic, M., & Chaumont, D. (2015). An evaluation of temperature and precipitation surface-based and reanalysis datasets for the Canadian Arctic, 1950–2010. *Atmosphere-Ocean*, *53*(3), 283–303. https://doi.org/10.1080/07055900.2015.1045825

Raphael, M. N., Marshall, G. J., Turner, J., Fogt, R. L., Schneider, D., Dixon, D. A., et al. (2016). The Amundsen Sea low variability, change, and impact on Antarctic climate. *Bulletin of the American Meteorological Society*, *97*(1), 111–121. https://doi.org/10.1175/bams-d-14-00018.1

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., et al. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, *108*(D14), 4407. https://doi.org/10.1029/2002jd002670

Sanz Rodrigo, J., Buchlin, J.-M., van Beeck, J., Lenaerts, J. T. M., & van den Broeke, M. R. (2013). Evaluation of the Antarctic surface wind climate from era reanalyses and RACMO2/ANT simulations based on automatic weather stations. *Climate Dynamics*, *40*(1), 353–376. https://doi.org/10.1007/s00382-012-1396-y

Sato, K., Inoue, J., Alexander, S. P., McFarquhar, G., & Yamazaki, A. (2018). Improved reanalysis and prediction of atmospheric fields over the Southern Ocean using campaign-based radiosonde observations. *Geophysical Research Letters*, *45*, 11,406–11,413. https://doi.org/10.1029/2018gl079037

Schneider, D. P., & Fogt, R. L. (2018). Artifacts in century-length atmospheric and coupled reanalyses over Antarctica due to historical data availability. *Geophysical Research Letters*, *45*, 964–973. https://doi.org/10.1002/2017gl076226

Screen, J. A., & Simmonds, I. (2011). Erroneous Arctic temperature trends in the ERA-40 reanalysis: A closer look. *Journal of Climate*, *24*(10), 2620–2627. https://doi.org/10.1175/2010jcli4054.1

Seefeldt, M. W., & Cassano, J. J. (2012). A description of the Ross Ice Shelf air stream (RAS) through the use of self-organizing maps (SOMs). *Journal of Geophysical Research*, *117*, D09112. https://doi.org/10.1029/2011jd016857

Seefeldt, M. W., Cassano, J. J., & Parish, T. R. (2007). Dominant regimes of the Ross Ice Shelf surface wind field during austral autumn 2005. *Journal of Applied Meteorology and Climatology*, *46*(11), 1933–1955. https://doi.org/10.1175/2007jamc1442.1

Silber, I., Verlinde, J., Cadeddu, M., Flynn, C. J., Vogelmann, A. M., & Eloranta, E. W. (2019). Antarctic cloud macrophysical, thermodynamic phase, and atmospheric inversion coupling properties at McMurdo stationpart II: Radiative impact during different synoptic regimes. *Journal of Geophysical Research: Atmospheres*, *124*, 1697–1719. https://doi.org/10.1029/2018jd029471

Simmons, A. J., & Poli, P. (2015). Arctic warming in ERA-Interim and other analyses. *Quarterly Journal of the Royal Meteorological Society*, *141*(689), 1147–1162. https://doi.org/10.1002/qj.2422

Stryhal, J., & Huth, R. (2017). Classifications of winter Euro-Atlantic circulation patterns: An intercomparison of five atmospheric reanalyses. *Journal of Climate*, *30*(19), 7847–7861. https://doi.org/10.1175/jcli-d-17-0059.1

Tastula, E. M., Vihma, T., Andreas, E. L., & Galperin, B. (2013). Validation of the diurnal cycles in atmospheric reanalyses over Antarctic sea ice. *Journal of Geophysical Research: Atmospheres*, *118*, 4194–4204. https://doi.org/10.1002/jgrd.50336

Theil, H. (1970). On the estimation of relationships involving qualitative variables. *American Journal of Sociology*, *76*(1), 103–154.

Thorne, P. W., & Vose, R. S. (2010). Reanalyses suitable for characterizing long-term trends. *Bulletin of the American Meteorological Society*, *91*(3), 353–362. https://doi.org/10.1175/2009bams2858.1

Torralba, V., Doblas-Reyes, F. J., & Gonzalez-Reviriego, N. (2017). Uncertainty in recent near-surface wind speed trends: A global reanalysis intercomparison. *Environmental Research Letters*, *12*(11), 9. https://doi.org/10.1088/1748-9326/aa8a58

Wang, Y., Zhou, D., Bunde, A., & Havlin, S. (2016). Testing reanalysis data sets in Antarctica: Trends, persistence properties, and trend significance. *Journal of Geophysical Research: Atmospheres*, *121*, 12,839–12,855. https://doi.org/10.1002/2016jd024864

Wen, C. H., Kumar, A., & Xue, Y. (2019). Uncertainties in reanalysis surface wind stress and their relationship with observing systems. *Climate Dynamics*, *52*(5-6), 3061–3078. https://doi.org/10.1007/s00382-018-4310-4

Wohland, J., Omrani, N. E., Witthaut, D., & Keenlyside, N. S. (2019). Inconsistent wind speed trends in current twentieth century reanalyses. *Journal of Geophysical Research: Atmospheres*, *124*, 1931–1940. https://doi.org/10.1029/2018jd030083

Zhang, Y. L., Wang, Y. T., Huai, B. J., Ding, M. H., & Sun, W. J. (2018). Skill of the two 20th century reanalyses in representing Antarctic near-surface air temperature. *International Journal of Climatology*, *38*(11), 4225–4238. https://doi.org/10.1002/joc.5563